



## King's Research Portal

DOI:

[10.1038/s41598-020-59256-0](https://doi.org/10.1038/s41598-020-59256-0)

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Powell, T. R., R. Duarte, R. R., Hotopf, M., Hatch, S. L., de Mulder, M., Breen, G. D., Lewis, C. M., & Nixon, D. F. (2020). The behavioral, cellular and immune mediators of HIV-1 acquisition: New insights from population genetics. *Scientific Reports*, 10(1), [3304 ]. <https://doi.org/10.1038/s41598-020-59256-0>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

OPEN

# The behavioral, cellular and immune mediators of HIV-1 acquisition: New insights from population genetics

Timothy R. Powell<sup>1,2,4\*</sup>, Rodrigo R. R. Duarte<sup>1,2,4</sup>, Matthew Hotopf<sup>3</sup>, Stephani L. Hatch<sup>3</sup>, Miguel de Mulder Rougvié<sup>2</sup>, Gerome D. Breen<sup>1</sup>, Cathryn M. Lewis<sup>1</sup> & Douglas F. Nixon<sup>1,2\*</sup>

Millions are exposed to the human immunodeficiency virus type 1 (HIV-1) every year, but not all acquire the virus, suggesting a potential role for host genetics in the moderation of HIV-1 acquisition. Here, we analyzed summary statistics from the largest genome-wide association study of HIV-1 acquisition to-date, consisting of 6,334 infected patients and 7,247 population controls, to advance our understanding of the genetic mechanisms implicated in this trait. We found that HIV-1 acquisition is polygenic and heritable, with SNP heritability estimates explaining 28–42% of the variance in this trait at a population level. Genetic correlations alongside UK Biobank data revealed associations with smoking, prospective memory and socioeconomic traits. Gene-level enrichment analysis identified EF-hand calcium binding domain 14 as a novel susceptibility gene for HIV-1 acquisition. We also observed that susceptibility variants for HIV-1 acquisition were significantly enriched for genes expressed in T-cells, but also in striatal and hippocampal neurons. Finally, we tested how polygenic risk scores for HIV-1 acquisition influence blood levels of 35 inflammatory markers in 406 HIV-1-negative individuals. We found that higher genetic risk for HIV-1 acquisition was associated with lower levels of C-C motif chemokine ligand 17. Our findings corroborate a complex model for HIV-1 acquisition, whereby susceptibility is partly heritable and moderated by specific behavioral, cellular and immunological parameters.

Around 38 million people currently live with the human immunodeficiency virus type 1 (HIV-1) worldwide<sup>1</sup>, and millions more are exposed to potential infection every year through sexual contact, vertical transmission, or via the parenteral route<sup>2,3</sup>. First-line prevention strategies against acquisition comprise of the use of condoms and pre-exposure prophylaxis (PrEP), or abstinence from drugs or sex<sup>4</sup>. However, epidemiological studies have identified varying degrees of susceptibility to HIV-1, suggesting that host genetics may play a role in moderating acquisition, which could be explored in the context of preventive strategies. For example, studies conducted prior to the development of antiretroviral therapy observed that less than a third of babies born from HIV-1-positive mothers acquire HIV-1<sup>5</sup> and, similarly, that a proportion of highly exposed individuals are resistant to infection<sup>6</sup>. Supporting this hypothesis, homozygosity of the  $\Delta 32$  mutation of the C-C chemokine receptor type 5 (CCR5) gene has been shown to protect against HIV-1 infection<sup>7–9</sup>, as the encoded protein is a co-receptor needed for viral entry. However, it remains unknown whether common genetic risk factors are also involved in host susceptibility to acquisition.

HIV-1 acquisition is a complex phenotype that consists of behavioral risk parameters and biological factors moderating viral entry and replication. A better understanding of both behavioral and biological factors influencing acquisition has the potential to improve our basic comprehension of acquisition, inform prevention strategies and clinical trials, and reduce social stigma. In this context, genome-wide association studies (GWAS) provide a powerful means to identify variants and risk mechanisms implicated in HIV-1 acquisition. No genome-wide significant polymorphisms have been robustly associated with HIV-1 acquisition to-date<sup>10–16</sup>.

<sup>1</sup>Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. <sup>2</sup>Division of Infectious Diseases, Weill Cornell Medicine, Cornell University, New York, NY, USA. <sup>3</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. <sup>4</sup>These authors contributed equally: Timothy R. Powell and Rodrigo R. R. Duarte. \*email: [timothy.1.powell@kcl.ac.uk](mailto:timothy.1.powell@kcl.ac.uk); [dnixon@med.cornell.edu](mailto:dnixon@med.cornell.edu)

However, population genetic methods have developed substantially in recent years, now allowing for powerful, biologically-informative analyses even in moderately-sized GWAS. For example, gene enrichment analyses can be applied to summary statistics to identify genes involved in risk, as well as cell types more likely mediating susceptibility. This is achieved by averaging association signals from multiple neighboring polymorphisms within protein-coding genes, and comparing lists of the resulting genes to the transcriptional profiles of mammalian cell types<sup>17,18</sup>. Methods like Linkage Disequilibrium Score Regression (LDSC)<sup>19</sup> and Linkage Disequilibrium Adjusted Kinships (LDAK)<sup>20</sup> further allow the estimation of SNP heritability based on GWAS results. Polygenic risk scores (PRS), in turn, can be used to explore genetic overlap between traits, and are useful in establishing the effect genetic predisposition has on biological parameters. PRS could even be used to model genetic risk for a trait (e.g. HIV-1 acquisition susceptibility) in a cohort of unaffected individuals (e.g. HIV-1-negative individuals). This, for example, could allow us to model the impact of genetic predisposition to HIV-1 acquisition on biological systems prior to HIV-1 infection, which could ultimately aid in the identification of new vaccine or drug development strategies. Here, we apply these modern population genetic methods to the largest GWAS of HIV-1 acquisition, and identify novel factors and complex traits associated with HIV-1 acquisition.

## Results

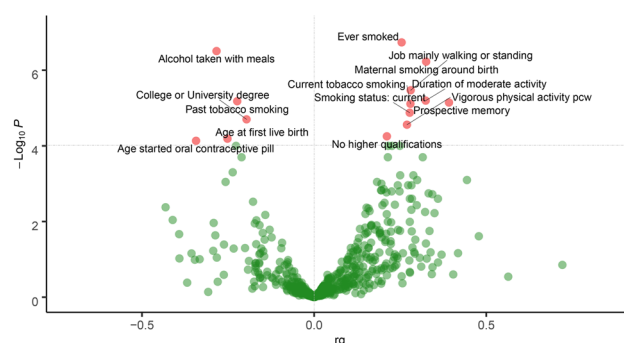
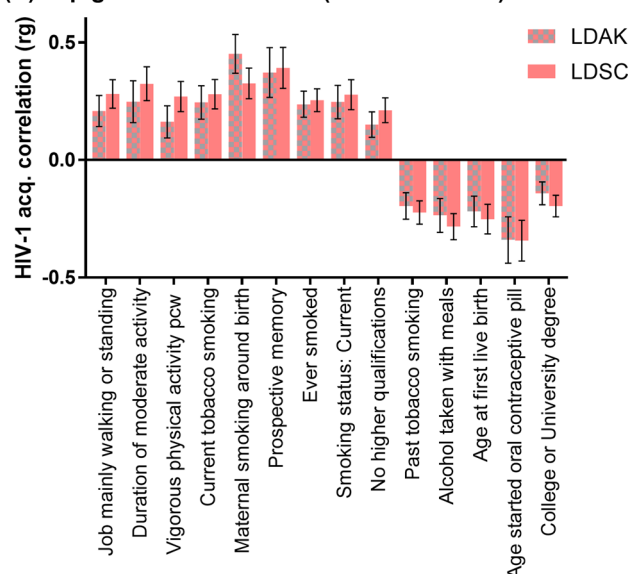
**HIV-1 acquisition is heritable and correlates with behavioral and socioeconomic traits.** The largest GWAS meta-analysis of HIV-1 acquisition to-date tested over 8 million common polymorphisms for association with HIV-1 acquisition in 6,334 infected patients and 7,247 population controls<sup>10</sup>. We estimated the SNP heritability (SNP  $h^2$ ) of HIV-1 acquisition by analyzing these GWAS results using LD Hub (LDSC-based)<sup>19</sup> and SumHer-GC<sup>20</sup> (based on the LDAK model, correcting for possible hidden population structure). We observed that  $\text{SNP } h^2 = 0.28 \pm 0.05$  (standard deviation; SD) under the LDSC model, suggesting that HIV-1 acquisition is a heritable trait, replicating a previous analysis of the same dataset<sup>21</sup>. Analysis with SumHer-GC showed a higher estimate of  $\text{SNP } h^2$  ( $0.42 \pm 0.08$ ), consistent with LDAK being able to capture a larger proportion of SNPs contributing to  $\text{SNP } h^2$ , relative to LDSC.

To better understand which behavioral parameters might be important in moderating HIV-1 acquisition, we performed genetic correlation analyses leveraging on GWAS data from 516 heritable traits assessed in the UK Biobank via LD Hub, which contains genetic association results from up to 488,377 individuals. This LDSC-based analysis revealed 9 positive correlations with HIV-1 acquisition, including *prospective memory*, ascertained using cognitive tests ( $rg = 0.39$ ,  $SE = 0.09$ ,  $P = 7.15 \times 10^{-6}$ ); lower levels of education, ascertained by having *no higher qualifications* ( $rg = 0.21$ ,  $SE = 0.05$ ,  $P = 5.59 \times 10^{-5}$ ); self-reported *smoking* ( $rg = 0.28$ ,  $SE = 0.06$ ,  $P = 1.33 \times 10^{-5}$ ); and self-reported *vigorous exercising* ( $rg = 0.27$ ,  $SE = 0.06$ ,  $P = 2.76 \times 10^{-5}$ ; Bonferroni adjusted  $P$  (for 516 traits)  $< 0.05$  for all; Fig. 1A; Supplemental Table 1). We also observed 5 negative associations with acquisition, including socioeconomic traits like *alcohol intake with meals* ( $rg = -0.28$ ,  $SE = 0.06$ ,  $P = 3.12 \times 10^{-7}$ ), *having a higher qualification* ( $rg = -0.20$ ,  $SE = 0.05$ ,  $P = 1.99 \times 10^{-5}$ ), and *age at which female participants had their first live birth* ( $rg = -0.25$ ,  $SE = 0.06$ ,  $P = 6.46 \times 10^{-5}$ ). We validated the Bonferroni-significant genetic correlations using SumHer-GC (Fig. 1B), and observed highly concordant estimates of genetic correlation calculated between the two methods (Pearson's  $r = 0.98$ ,  $P = 2.59 \times 10^{-9}$ ).

**EFCAB14 as a novel susceptibility gene for HIV-1 acquisition.** To perform our gene-level enrichment analysis, gene-level statistics and weighted p-values were generated from GWAS summary statistics using MAGMA<sup>22</sup>, adjusting associations for gene size, single nucleotide polymorphism (SNP) density and linkage disequilibrium. This analysis revealed a contribution of several protein-coding genes to HIV-1 acquisition, with gene-level Q-Q plots highlighting the degree of polygenicity observed, and the abundance of contributing genes relative to an expected normal distribution (Fig. 2A; Supplemental Table 2). We identified a novel gene involved in HIV-1 acquisition, the EF-hand calcium binding domain 14 (*EFCAB14*), on chromosome 1p33 ( $Z\text{-score} = 4.56$ , enrichment  $P = 2.56 \times 10^{-6}$ , Bonferroni corrected  $P = 4.7 \times 10^{-2}$ ), which is expressed ubiquitously across tissues, according to The Genotype-Tissue Expression (GTEx) project<sup>23</sup>. The highest association signal at this locus is the non-coding variant rs8851 ( $P = 5.57 \times 10^{-7}$ ; Fig. 2B), which is also an expression quantitative trait loci (eQTL) for *EFCAB14*. The risk (G-) allele of rs8851 is associated with lower expression of *EFCAB14* in multiple tissues, including whole blood, skin, adipose tissue, the cerebellum and arteries ( $P < 1 \times 10^{-3}$  for all)<sup>23</sup>. These findings suggest that genetic risk for HIV-1 acquisition at this locus is conferred via reduced expression of *EFCAB14*, and not by altered expression of neighboring genes such as the ATP synthase mitochondrial F1 complex assembly factor 1 (*ATPAF1*) or the testis expressed 38 (*TEX38*) genes (Fig. 2C). Outside of chromosome 1p33, the top gene-level association signal was on chromosome 15q22.31, at the ubiquitin specific peptidase 3 gene (*USP3*), although this was not significant after multiple testing correction (enrichment  $P = 4.96 \times 10^{-6}$ , Bonferroni corrected  $P > 0.05$ ; Fig. 2D).

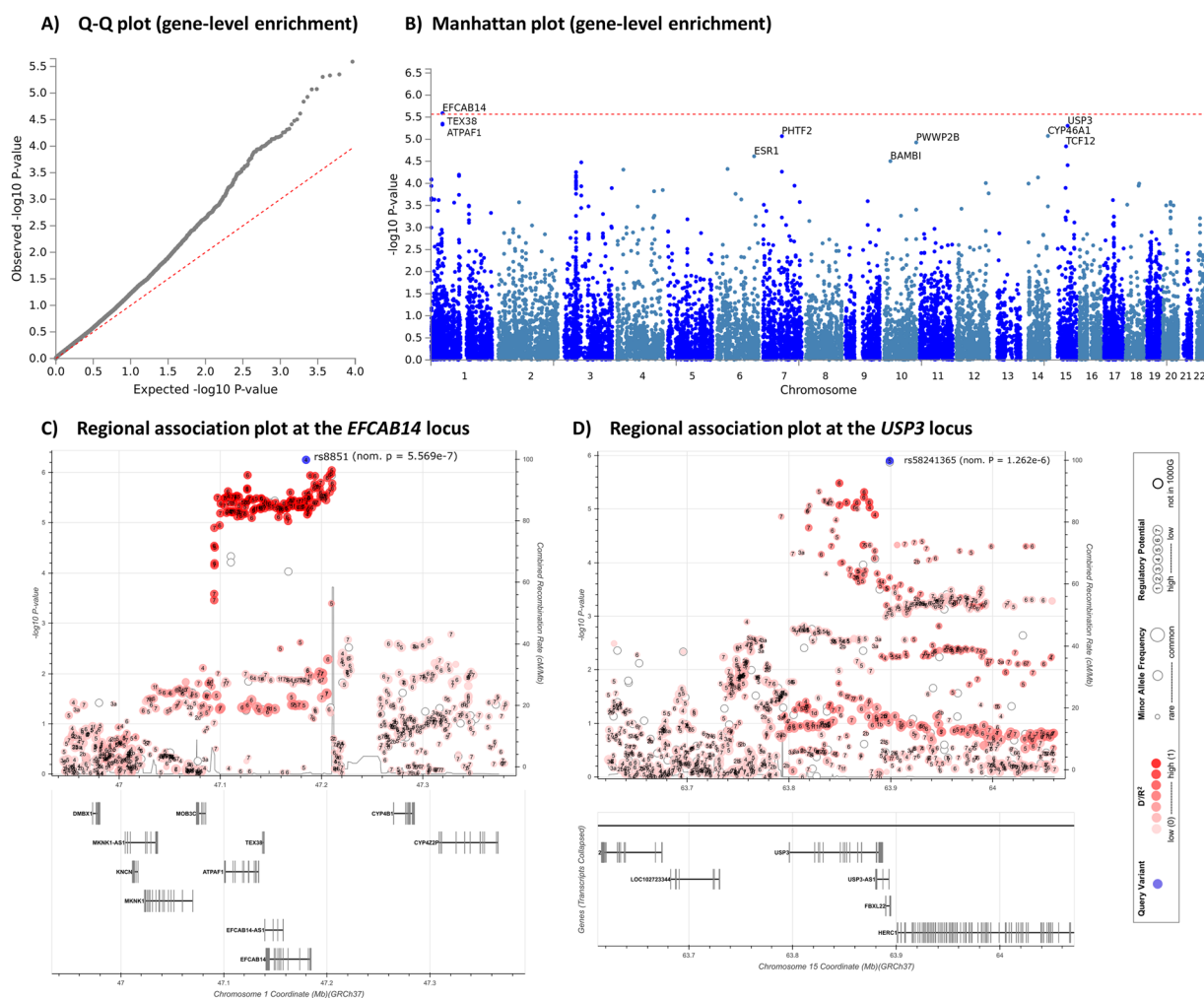
**HIV-1 acquisition susceptibility variants are enriched within genes expressed in T-cells and striatal neurons.** We aimed to investigate the cellular basis for the biological and behavioral parameters implicated in HIV-1 acquisition, and therefore investigated the cell types enriched for variants associated with this trait. Two independent gene-set enrichment analyses using FUMA<sup>24</sup> showed that cells likely mediating acquisition included T-cells ( $P = 8.54 \times 10^{-4}$ ) and, independently, neurons from the striatum, hippocampus and globus pallidus under a false discovery rate of 10% (top association signals, respectively:  $P = 1.22 \times 10^{-4}$ ,  $2.49 \times 10^{-4}$ , and  $6.23 \times 10^{-4}$ ;  $q < 0.10$  for all; Fig. 3A,B, Supplemental Tables 3 and 4).

**Polygenic risk score for HIV-1 acquisition is negatively associated with circulating CCL17 levels.** We utilized findings from the GWAS performed by McLaren and colleagues (2013) to calculate PRS for HIV-1 acquisition in an unrelated cohort of HIV-1-negative individuals, to measure how acquisition

**(A) Genetic correlations with HIV-1 acq. (LDSC)****(B) Top genetic correlations (LDSC vs LDAK)**

**Figure 1.** Genetic correlations between HIV-1 acquisition susceptibility and traits tested within the UK Biobank. **(A)** Correlations performed within LD Hub using the LDSC model to determine genetic correlations between HIV-1 acquisition and 516 heritable traits. The Bonferroni-significant correlations ( $rg$ ) are displayed in red and delimited by the horizontal line ( $P = 0.05/516 \text{ traits} = 8.38 \times 10^{-5}$ ). **(B)** Validation of the Bonferroni-significant findings using SumHer-GC, which is based on the LDAK model, adjusting for genomic control. The correlation values observed for the 14 Bonferroni-significant traits associated with HIV-1 acquisition in the LDSC method were highly concordant with results using the LDAK method (Pearson's  $r = 0.98$ ,  $P = 2.59 \times 10^{-9}$ ).

predisposition correlated with levels of 35 blood-based inflammatory markers. Our rationale was to better understand how genetic susceptibility expresses itself in the pre-exposed immune system, and in particular how it affects inflammatory cytokines, which are molecules that can be modified via pharmacological intervention, and are thought to be key moderators of HIV-1 infection<sup>25,26</sup>. A preliminary analysis utilized all polymorphisms associated with HIV-1 acquisition under a  $P$  association threshold ( $P_T$ ) = 0.5 to calculate the PRS, and to test its correlation with the levels of 35 inflammatory markers. This analysis showed that HIV-1 acquisition was significantly and specifically associated with lower levels of the chemokine CCL17 ( $\beta = -1644.32$ , standard error (SE) = 496.62,  $P = 1.00 \times 10^{-3}$ , Bonferroni corrected  $P$  (for 35 tests) =  $3.50 \times 10^{-2}$ , variance explained = 2.74%; Fig. 3C). A sensitivity analysis revealed that PRS at every tested  $P_T$  also significantly predicted levels of this chemokine, with highest significance at  $P_T = 0.061$  ( $\beta = -800.52$ , SE = 180.07,  $P = 1.14 \times 10^{-5}$ , corrected  $P = 5.11 \times 10^{-4}$ , variance explained = 4.84%; Fig. 3D). Importantly, this effect was additionally observed when considering individuals of European-only ancestry in the cohort ( $\beta = -794.27$ , SE = 190.81  $P = 4.21 \times 10^{-5}$ ), matching the ethnicity of the individuals in the base GWAS used to construct the PRS, and after removing the major histocompatibility complex from the PRS calculation ( $\beta = -526.42$ , SE = 158.08,  $P = 9.50 \times 10^{-4}$ ). The inverse correlation observed suggests that individuals that have a higher genetic predisposition to HIV-1 acquisition are more likely to have lower blood levels of CCL17 (Fig. 3E).



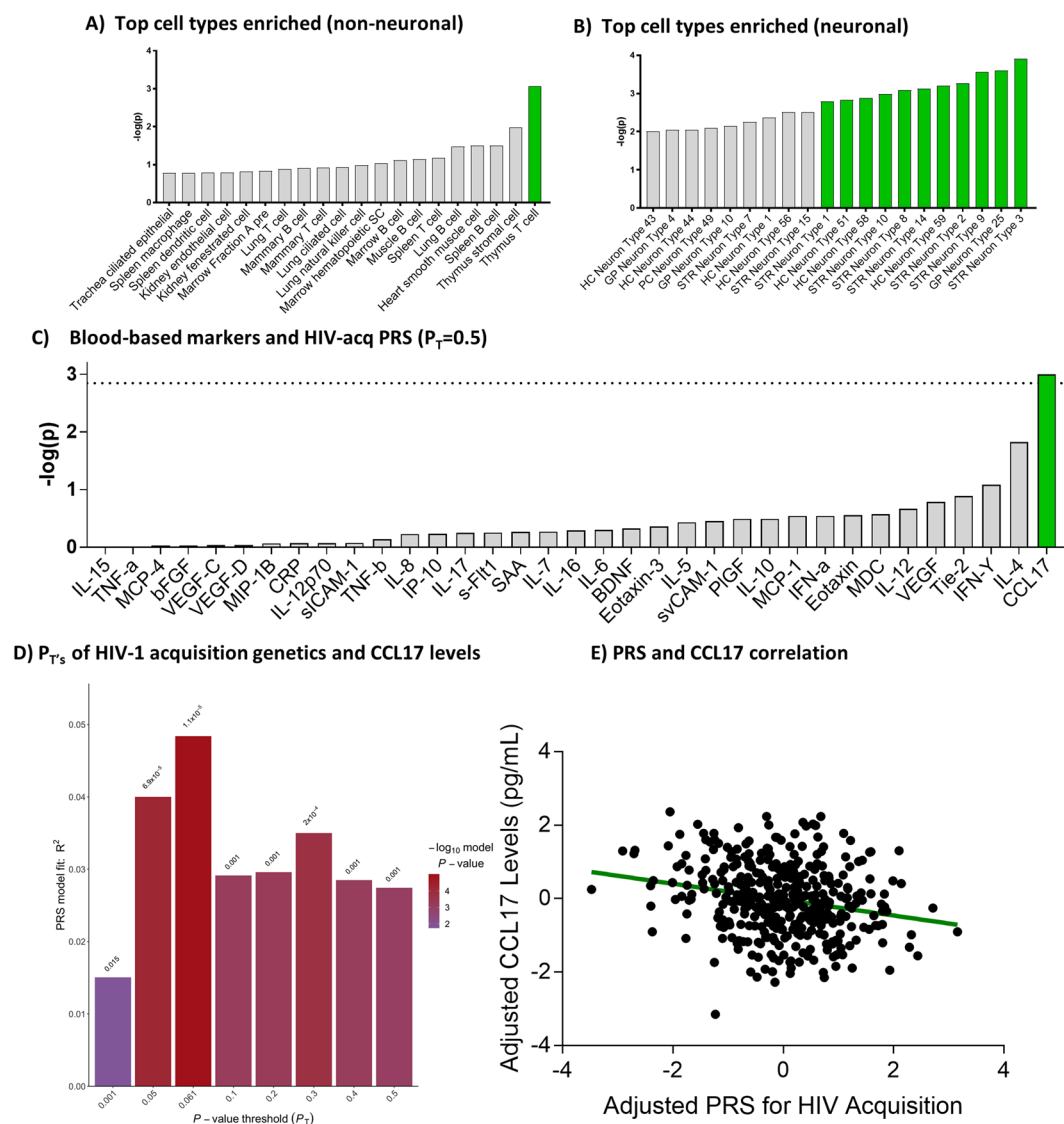
**Figure 2.** Gene-level enrichment analysis of the HIV-1 acquisition GWAS summary statistics identified a Bonferroni-significant susceptibility gene. **(A)** The quantile-quantile (Q-Q) plot shows the high number of observed genes associated with HIV-1 acquisition, compared to the number of expected genes assuming a normal distribution (dotted red line). Generated using FUMA<sup>17</sup>. **(B)** Gene-level Manhattan plot showing the novel susceptibility gene for HIV-1 acquisition on chromosome 1p33, *EFCAB14*. Generated using FUMA<sup>17</sup>. **(C)** Regional association plot demonstrating high linkage disequilibrium at the *EFCAB14* locus. Generated using LDassoc<sup>51</sup>. **(D)** Regional association plot at the second highest association signal with HIV-1 acquisition outside of chromosome 1. Generated using LDassoc<sup>51</sup>.

## Discussion

HIV-1 acquisition consists of behavioral risk parameters moderating exposure as well as biological factors controlling viral entry and replication. However, the genetic aspects of HIV-1 acquisition (outside of the CCR5  $\Delta 32$  mutation) have been understudied, in part because no robust genome-wide significant variants were found in early studies<sup>10–16</sup>, which led to the premature assumption that acquisition was not substantially moderated by common genetic variants. The GWAS analyzed here, by McLaren and colleagues (2013), did not identify genome-wide significant polymorphisms associated with HIV-1 acquisition (after correcting association signals for frailty bias), but population genetic methods have advanced considerably in recent years, now allowing for powerful inferences about genetic traits using GWAS summary statistics<sup>19,22,24,27</sup>, even in moderately powered studies. For instance, we calculated heritability estimates of HIV-1 acquisition using cutting edge methods like LDSC and LDK, which would otherwise be challenging using traditional twin and family methods. The level of SNP  $h^2$  observed for acquisition (LDSC:  $0.28 \pm 0.05$ ; LDK:  $0.42 \pm 0.08$ ) was greater or comparable to that of traits considered highly heritable, such as body mass index (LDSC:  $0.09 \pm 0.01$ ; LDK:  $0.33 \pm 0.03$ ), height (LDSC:  $0.20 \pm 0.02$ ; LDK:  $0.46 \pm 0.04$ ), and schizophrenia (LDSC:  $0.19 \pm 0.01$ ; LDK:  $0.42 \pm 0.02$ )<sup>20</sup>. Overall, these results highlight the contribution of common variants to HIV-1 acquisition risk, showing this is a heritable trait.

To understand the underlying genetic factors associated with HIV-1 acquisition, we performed genetic correlation analyses in LD Hub, which leverages on data from ~500,000 individuals from the UK Biobank, to investigate how acquisition genetics correlates with heritable traits assessed in this large population sample. We observed





**Figure 3.** HIV-1 acquisition genetics significantly overlaps with the expression profile of T-cells and neurons, and correlates with blood levels of CCL17. **(A)** Top 20 enrichment signals observed for murine non-neuronal cell types. **(B)** Top 20 enrichment signals observed for murine neuronal cell types. Green indicates significance under a false discovery rate of 10%. **(C)** Plot showing  $-\log(p)$  of the association between 35 markers and a polygenic risk score for HIV-1 acquisition. The dashed line indicates Bonferroni significance. **(D)** Sensitivity analysis revealed that PRS at every  $P_T$  significantly correlated with CCL17 levels in blood, with the highest significance at  $P_T=0.061$ . **(E)** Correlation between PRS for HIV-1 acquisition ( $P_T=0.061$ ) adjusted for seven population dimensions, and CCL17 levels adjusted for age, gender, BMI, ethnicity, and smoking status.

genetic correlations between HIV-1 acquisition and heritable phenotypes associated with socio-economic factors, corroborating previous epidemiological work, and further highlighting the need for prevention strategies tailored to individuals who most need it<sup>28</sup>. We further validated the genetic correlations using the independent SumHer-GC method, supporting these results.

Our results also validate and expand the current understanding of the biological basis of HIV-1 acquisition. In a preliminary enrichment analysis that aimed to identify the cell types that mediate HIV-1 acquisition susceptibility throughout the body, we observed that polymorphisms implicated in acquisition were enriched for genes expressed in T-cells, which are the main targets for HIV-1 replication<sup>29</sup>. We further tested the enrichment of HIV-1 variants for genes expressed across a range of neural cell types, since these cells mediate behavior and could explain certain genetic correlations observed with HIV-1 acquisition. We observed a significant enrichment for striatal and hippocampal neurons in association with HIV-1 acquisition, which is particularly striking considering they are brain areas implicated in the regulation of reward and pleasure<sup>30,31</sup>. Alternatively, these cell types may represent those which harbor HIV-1 and most effectively hide it from the immune system, propagating a sustained infection. Furthermore, the gene-level enrichment analysis identified *EFCAB14* as a susceptibility gene for HIV-1 acquisition, on chromosome 1p33. This gene is ubiquitously expressed in the body, and the risk

allele of rs8851 is known to reduce expression of *EFCAB14* across multiple tissues. Proteins containing EF-Hand Calcium Binding domains in general are implicated in functions ranging from intracellular calcium buffering, signal transduction and muscle contraction<sup>32</sup>, but future studies are warranted to investigate the function of *EFCAB14* specifically, particularly in the context of HIV-1 acquisition.

Another emerging method in population genetics is polygenic risk scoring<sup>33,34</sup>. We modelled how genetic risk for HIV-1 acquisition expresses itself in the pre-exposed immune profile using a cohort of HIV-1 negative individuals. By considering genetic risk as a continuous trait in a population setting, we can more powerfully determine the influences of the genetic risk signal on innate biological systems such as inflammatory marker expression, without confounders (e.g. drug use, other infections) more commonly associated with individuals from high-risk groups. Moreover, previous studies that have compared high-risk individuals who do not acquire HIV-1, with those that do, are likely confounded by the fact that HIV-1 has an influence on the immune system and inflammatory profile of the individual, which may not correspond to the pre-exposed immune profile associated with risk or resilience. In particular, we studied how genetic predisposition to HIV-1 acquisition affects inflammatory cytokines, which are immune messengers that are relatively easy to assay, can be modified via pharmacological intervention, and are thought to be key moderators of HIV-1 infection<sup>25,26</sup>. We observed that PRS for HIV-1 acquisition inversely correlated with CCL17 levels in the blood of HIV-1 negative individuals, suggesting that levels of this chemokine should be considered in clinical trials for biomarker, drug and vaccine development. CCL17 is known to regulate the development and maturation of T-cells in the thymus, as well as their trafficking during inflammation<sup>35–37</sup>. Neutralization of this chemokine by antibody treatment has been shown to block the recruitment of T-cells in the lung (ameliorating respiratory allergy)<sup>38</sup>. We hypothesize that increased CCL17 levels may increase the influx of inflammatory cells, which could help eliminate HIV-1-infected cells before the establishment of a systemic infection. However, CCL17 levels may represent only one of many biological mechanisms implicated in, or co-occurring with, HIV-1 acquisition susceptibility, and further research is needed to better understand this relationship.

Despite the insights provided, our study has limitations, including the modest cohort size of European-only individuals in the GWAS analyzed. Analyses of larger cohorts from different ancestry groups and well-characterized infection routes have the potential to improve our understanding of HIV-1 acquisition, by improving the identification of specific SNPs and genes involved. Future well-powered GWAS of HIV-1 acquisition comparing high risk individuals who are infected versus those who are not will more likely tease apart the biological risk mechanisms implicated in viral resilience from the behavioral risk factors. This could be studied in areas where the risk for HIV-1 is already relatively high in the general population (e.g. South Africa), and where viral entry and replication may represent more important components of acquisition than risk behaviors. Furthermore, although we validated our estimates of heritability and genetic correlations by using two independent methods, additional cohorts are needed to replicate our findings, including those obtained with the PRS analysis. Moreover, our genetic correlation analysis relies on self-report information from the UK Biobank, where individuals (aged 40+) are probably older than the average age of individuals diagnosed with HIV-1. Consequently, behaviors relevant to HIV-1 acquisition that may be more common in a younger cohort (e.g. drug use), could have been absent (or not reported) in the UK Biobank. Where we do find genetic correlations with HIV-1 risk, we cannot currently infer cause and effect, but this is something which may be achievable in the future via a Mendelian randomization design, once larger GWAS are able to detect robust genome-wide significant predictors of HIV-1 acquisition. Finally, the cytokine panel we investigated in association with PRS for acquisition was limited to 35 proteins, and it is possible that other inflammatory markers which we did not assess may be more relevant to this trait.

To conclude, our results show that HIV-1 acquisition genetics impinges upon behavioral, cellular, and immune factors. By leveraging on modern population genetic methods, our work provides a novel framework to study HIV-1 acquisition as a complex phenotype, advancing our understanding of the underlying risk factors. Our results suggest that in addition to environmental risk factors, there is a polygenic component to HIV-1 acquisition that should be explored further in clinical studies. In addition, our work supports the investigation of future intervention strategies surrounding education and smoking behavior. In particular, there needs to be studies investigating whether smoking simply represents a proxy for unhealthy behaviors, or whether it actively influences biological processes linked to HIV-1 acquisition. Similarly, CCL17 and *EFCAB14* should also be investigated with respect to their potential role as biomarkers for HIV-1 acquisition and as drug targets. As demonstrated here, GWAS can be very informative even when analyzing moderately sized cohorts, but its true potential to unveil the host genetic mechanisms influencing HIV-1 acquisition will likely only be unlocked with the creation of collaborative initiatives and analyses of larger cohorts.

## Material and Methods

**GWAS summary statistics.** We obtained summary statistics from the largest HIV-1 acquisition GWAS meta-analysis to-date, performed by McLaren and colleagues, which analyzed 6,334 HIV-1-positive individuals and 7,247 HIV-1-negative population controls of European ancestry, collected in Europe, North America, Australia and Africa<sup>10</sup>. The HIV-1-positive group consisted of individuals who acquired HIV-1 by sexual contact (homosexual and heterosexual,  $N = 3,311$ ), via the parenteral route (injection drug use and transfusion,  $N = 1,046$ ), or by unknown means ( $N = 2,086$ ). The authors of this GWAS found no association of genotype with specific means of acquisition. We excluded variants not genotyped across all studies in the meta-analysis, those with a minor allele frequency  $< 0.01$ , and those located within the major histocompatibility complex (chromosome 6, 26–34 Mb), due to its complex linkage disequilibrium structure.

**Heritability estimates and genetic correlations.** We applied LD score regression in LD Hub v1.9.1<sup>19</sup> to estimate the heritability of HIV-1 acquisition and to identify genetically associated traits, based on 516 heritable

phenotypes assessed in the UK Biobank<sup>39</sup>. The UK Biobank phenotypically characterized and genotyped 488,377 participants aged between 40–69 years, enabling LD Hub to test for genetic correlations between HIV-1 acquisition and traits of interest (e.g. disease states, behaviors and socioeconomic traits)<sup>40</sup>. Results were plotted in R<sup>41</sup> using the “EnhancedVolcano” library<sup>42</sup>. Estimation of SNP  $h^2$  and genetic correlations were also performed using SumHer-GC, from the LDAK package<sup>20</sup>, with the genomic control feature enabled, to control for potential hidden population structure in the data, even though the authors of the original GWAS did not identify genomic inflation (see McLaren and colleagues (2013)<sup>10</sup>). In the LDSC model, all SNPs are assumed to contribute equally to the phenotype, whereas in the LDAK model (which is expected to capture a larger proportion of heritability), a SNP with higher frequency is expected to contribute more towards heritability than one with lower frequency in the population, and a SNP in a region of low linkage disequilibrium contributes more than one in a region of high linkage disequilibrium.

**Gene-level and gene set enrichment analyses.** We used the Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) to identify genes and cell types associated with HIV-1 acquisition genetics, as described elsewhere<sup>17,18</sup>. Briefly, gene-level statistics and weighted p-values were generated from GWAS summary statistics, adjusting gene-level associations for gene size, SNP density and LD. Input SNPs were mapped to 18,439 protein coding genes using a 10 kb upstream and downstream window, and a Bonferroni cut-off was applied to determine significance ( $P$  cut-off =  $0.05/18,439$  protein-coding genes =  $2.71 \times 10^{-6}$ )<sup>19</sup>. Next, the gene-set enrichment analysis used two single-cell RNA-sequencing datasets available in FUMA to identify cell types enriched for genes associated with HIV-1 acquisition under the false discovery rate of 10% ( $q < 0.10$ ). Specifically, to identify particular tissues or organs involved in HIV-1 acquisition, we tested the transcriptional profile of 75 murine cell types (TabulaMuris\_droplet\_all) for enrichment with HIV-1 acquisition susceptibility genes. This dataset is a comprehensive collection of well-curated single cell transcriptome data from *Mus musculus*, containing information from 100,605 cells from 20 organs and tissues<sup>43</sup>. To identify specific neural populations that could drive the behaviors observed in the genetic correlations, we tested the transcriptional profile of 565 neuronal cell types (DropViz\_all\_level2) for enrichment with HIV-1 acquisition susceptibility genes. This dataset contains the transcriptional signature of 690,000 cells from the mouse brain, which have been previously used to investigate the cellular mechanisms of behavior<sup>44</sup>.

**The South East London Community Health Study.** We aimed to assess how HIV-1 acquisition risk might be moderated by an individual's immune profile prior to infection, and so we tested how polygenic risk for HIV-1 acquisition correlated with the expression of 35 inflammatory markers. We studied HIV-1-negative population controls from the South East London Community Health Study (SELCoH)<sup>45–47</sup>, where HIV-1-negative status was determined based on self-report. For further details on the full SELCoH study, please see Hatch *et al.*<sup>43</sup>. SELCoH aimed to investigate mental and physical health in the general population in London, UK. The sub-sample analyzed in our study consisted of 406 individuals for which both inflammatory and genetic data were available. The SELCoH study received approval from King's College London research ethics committee, reference PNM/12/13-152. Informed written consent was obtained from all participants at the time of sample collection. The 35 inflammatory markers represent adequately expressed cytokines from an initial panel of 42, which were originally assayed in relation to major depression risk, as described previously<sup>48</sup>. The mean age of our sample was  $48.7 \pm 15.1$  (standard deviation), with a mean body mass index of  $27.3 \pm 5.5$ . The cohort is representative of the source population and consisted of 45.3% males; 20.9% current smokers and 40.4% ex-smokers; and 56.8% White British, 8.4% Black Caribbean, 10.9% Black African, 14.6% White Other, 6.2% Non-White Other, and 3.2% Mixed. Participants received detailed and repeated phenotypic assessments as part of three separate phases. The first phase aimed to assess common physical and mental disorders in South East London; the second, to examine the roles of historical social context and policy in shaping patterns of health inequalities; and the third, to collect biological specimens from a subset of participants, including blood for serum separation and DNA for genotyping. Serum and DNA were extracted and stored at  $-80^\circ\text{C}$  until use, as described previously<sup>47,48</sup>.

**Quantification and analysis of cytokines.** Serum levels (pg/mL) of 35 blood-based markers were assessed in blood samples from the SELCoH cohort using multiplex ELISA-based technology provided by the Meso Scale Discovery Biomarker kits, as described previously<sup>48</sup>. For each marker, we adjusted for the effects of assay run, age, gender, body mass index (BMI), smoking (never, current, former) and ethnicity, by taking standardized residuals (z-scores). Z-scores were then used in downstream PRS analyses.

**DNA genotyping.** DNA samples from the SELCoH cohort were sent to the Affymetrix Research Services Laboratory in Santa Clara, California, USA. Genotyping was assayed using the UK Biobank Axiom Array (r3) which comprises of 820,967 genetic markers (Affymetrix, California, United States). Genotype data was put through quality control measures as described previously, and used to construct polygenic risk scores<sup>47</sup>.

**Polygenic risk scores.** Individualized polygenic risk scores (PRS) within the SELCoH sample were calculated using PRSice-2, a PRS quantification software<sup>33,34</sup>. This pipeline uses summary statistics from a base GWAS to generate individualized PRS in a target dataset. Briefly, the number of risk alleles in the base dataset are multiplied by SNPs' effect sizes to generate individualized PRS in the target dataset. PRSice-2 clumps SNPs in the genotype files of the target dataset and removes those in high LD to avoid polygenic score inflation. For our initial screen, we constructed a PRS using all SNPs in the GWAS with  $P < 0.5$ . We tested whether this score predicted adjusted inflammatory marker levels in PRSice-2, covarying for seven population dimensions generated using principal component analysis applied to LD-pruned SNPs in PLINK 1.9<sup>49</sup>. Our relaxed p-value threshold ( $P_T$ ) was initially selected based on earlier studies which revealed that highly polygenic phenotypes in moderately powered GWAS are better captured using a relaxed cut-off<sup>50</sup>. We subsequently performed sensitivity analyses for



significant associations using a wide range of p-value thresholds, from  $p = 0.001$  to  $p = 0.5$ , with 0.001 increments. This allowed us to determine the optimal  $P_T$  that explained the most variance. Furthermore, as part of a sensitivity test, once an optimal polygenic predictor was established, we tested its ability to predict inflammatory markers: (i) once the MHC region was excluded, (ii) when applied to individuals of European-only ancestry. This was performed to confirm that the result was not driven by confounding effects related to the complex structure of the MHC region, or to the genetic admixture in the SELCoH sample.

## Data availability

Summary statistics were made available to us upon request to McLaren and colleagues<sup>10</sup>. Due to ethical restrictions, the SELCoH data is not publicly available. Details on the SELCoH sample and requests to access phenotype and inflammation data can be made at <http://www.slam.nhs.uk/research/selcoh/selcoh-projects>. Access to SELCoH genetic data requires local approval via the NIHR Bioresource (bioresource@kcl.ac.uk). Genetic correlations were performed using LD Hub and UK Biobank data, which are open access (<http://ldsc.broadinstitute.org/ldhub/>).

Received: 30 August 2019; Accepted: 13 January 2020;

Published online: 24 February 2020

## References

- World Health Organisation. Global Health Observatory (GHO) Data, <https://www.who.int/gho/hiv/en/>. Date of access: 03/02/2020 (2018).
- Logie, C. H., James, L., Tharao, W. & Loutfy, M. R. HIV, gender, race, sexual orientation, and sex work: a qualitative study of intersectional stigma experienced by HIV-positive women in Ontario, Canada. *PLoS Med.* **8**, e1001124, <https://doi.org/10.1371/journal.pmed.1001124> (2011).
- Freeman, R. *et al.* Critical race theory as a tool for understanding poor engagement along the HIV care continuum among African American/Black and Hispanic persons living with HIV in the United States: a qualitative exploration. *Int. J. Equity Health* **16**, 54, <https://doi.org/10.1186/s12939-017-0549-3> (2017).
- Woodson, E. *et al.* HIV transmission in discordant couples in Africa in the context of antiretroviral therapy availability. *AIDS* **32**, 1613–1623, <https://doi.org/10.1097/QAD.0000000000001871> (2018).
- The Working Group on Mother-To-Child Transmission of HIV. Rates of mother-to-child transmission of HIV-1 in Africa, America, and Europe: results from 13 perinatal studies. The Working Group on Mother-To-Child Transmission of HIV. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology: Official Publication of the International Retrovirology Association* **8**, 506–510 (1995).
- Fowke, K. R. *et al.* Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. *Lancet* **348**, 1347–1351, [https://doi.org/10.1016/S0140-6736\(95\)12269-2](https://doi.org/10.1016/S0140-6736(95)12269-2) (1996).
- Marmor, M., Hertzmark, K., Thomas, S. M., Halkitis, P. N. & Vogler, M. Resistance to HIV infection. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* **83**, 5–17, <https://doi.org/10.1007/s11524-005-9003-8> (2006).
- Shea, P. R., Shianna, K. V., Carrington, M. & Goldstein, D. B. Host genetics of HIV acquisition and viral control. *Annual Review of Medicine* **64**, 203–217, <https://doi.org/10.1146/annurev-med-052511-135400> (2013).
- Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–377, [https://doi.org/10.1016/S0092-8674\(00\)80110-5](https://doi.org/10.1016/S0092-8674(00)80110-5) (1996).
- McLaren, P. J. *et al.* Association Study of Common Genetic Variants and HIV-1 Acquisition in 6,300 Infected Cases and 7,200 Controls. *Plos Pathogens* **9**, e1003515, <https://doi.org/10.1371/journal.ppat.1003515> (2013).
- Joubert, B. R. *et al.* A whole genome association study of mother-to-child transmission of HIV in Malawi. *Genome Med.* **2**, 17, <https://doi.org/10.1186/gm138> (2010).
- Petrovski, S. *et al.* Common human genetic variants and HIV-1 susceptibility: a genome-wide survey in a homogeneous African population. *AIDS* **25**, 513–518, <https://doi.org/10.1097/QAD.0b013e328343817b> (2011).
- Luo, M. *et al.* A Genetic Polymorphism of FREM1 is Associated with Resistance against HIV Infection in the Pumwani Sex Worker Cohort. *Journal of Virology* **86**, 11899–11905, <https://doi.org/10.1128/JVI.01499-12> (2012).
- NAID Center for HIV/AIDS Vaccine Immunology. *et al.* A genome-wide association study of resistance to HIV infection in highly exposed uninfected individuals with hemophilia A. *Human Molecular Genetics* **22**, 1903–1910, <https://doi.org/10.1093/hmg/ddt033> (2013).
- Johnson, E. O. *et al.* Novel Genetic Locus Implicated for HIV-1 Acquisition with Putative Regulatory Links to HIV Replication and Infectivity: A Genome-Wide Association Study. *Plos one* **10**, e0118149, <https://doi.org/10.1371/journal.pone.0118149> (2015).
- Lingappa, J. R. *et al.* Genomewide association study for determinants of HIV-1 acquisition and viral set point in HIV-1 serodiscordant couples with quantified virus exposure. *Plos one* **6**, e28632–e28632, <https://doi.org/10.1371/journal.pone.0028632> (2011).
- Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826, <https://doi.org/10.1038/s41467-017-01261-5> (2017).
- Schijven, D. *et al.* Comprehensive pathway analyses of schizophrenia risk loci point to dysfunctional postsynaptic signaling. *Schizophr Res* **199**, 195–202 (2018).
- Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279, <https://doi.org/10.1093/bioinformatics/btw613> (2017).
- Speed, D. & Balding, D. J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nature Genetics* **51**, 277–284, <https://doi.org/10.1038/s41588-018-0279-5> (2019).
- Power, R. A. *et al.* A genome-wide polygenic approach to HIV uncovers link to inflammatory bowel disease and identifies potential novel genetic variants. *bioRxiv*, 145383, <https://doi.org/10.1101/145383> (2017).
- de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *Plos Computational Biology* **11**, e1004219, <https://doi.org/10.1371/journal.pcbi.1004219> (2015).
- GTEX Consortium *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204 (2017).
- Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* **8**, 1826, <https://doi.org/10.1038/s41467-017-01261-5> (2017).
- Grobler, A. *et al.* Genital Inflammation and the Risk of HIV Acquisition in Women. *Clinical Infectious Diseases* **61**, 260–269, <https://doi.org/10.1093/cid/civ298> (2015).
- Zidek, Z., Anzenbacher, P. & Kmonícková, E. Current status and challenges of cytokine pharmacology. *British Journal of Pharmacology* **157**, 342–361, <https://doi.org/10.1111/j.1476-5381.2009.00206.x> (2009).

27. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295, <https://doi.org/10.1038/ng.3211> (2015).
28. Bunyasi, E. W. & Coetzee, D. J. Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of South Africa. *BMJ Open* **7**, e016232, <https://doi.org/10.1136/bmjopen-2017-016232> (2017).
29. Lee, B., Sharron, M., Montaner, L. J., Weissman, D. & Doms, R. W. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5215–5220 (1999).
30. Enkavi, A. Z. *et al.* Evidence for hippocampal dependence of value-based decisions. *Scientific Reports* **7**, 17738–17738, <https://doi.org/10.1038/s41598-017-18015-4> (2017).
31. Goulet-Kennedy, J., Labbe, S. & Fecteau, S. The involvement of the striatum in decision making. *Dialogues in Clinical Neuroscience* **18**, 55–63 (2016).
32. Lewit-Bentley, A. & Réty, S. EF-hand calcium-binding proteins. *Current Opinion in Structural Biology* **10**, 637–643, [https://doi.org/10.1016/S0959-440X\(00\)00142-1](https://doi.org/10.1016/S0959-440X(00)00142-1) (2000).
33. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: Polygenic Risk Score software. *Bioinformatics* **31**, 1466–1468, <https://doi.org/10.1093/bioinformatics/btu848> (2015).
34. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *GigaScience* **8**, <https://doi.org/10.1093/gigascience/giz082> (2019).
35. Bonner, K., Pease, J. E., Corrigan, C. J., Clark, P. & Kay, A. B. CCL17/thymus and activation-regulated chemokine induces calcitonin gene-related peptide in human airway epithelial cells through CCR4. *Journal of Allergy and Clinical Immunology* **132**, 942–950.e943, <https://doi.org/10.1016/j.jaci.2013.04.015> (2013).
36. Teran, L. M., Ramirez-Jimenez, F., Soid-Raggi, G. & Velazquez, J. R. Interleukin 16 and CCL17/thymus and activation-regulated chemokine in patients with aspirin-exacerbated respiratory disease. *Annals of Allergy, Asthma & Immunology* **118**, 191–196, <https://doi.org/10.1016/j.anai.2016.11.004> (2017).
37. Shimada, Y., Takehara, K. & Sato, S. Both Th2 and Th1 chemokines (TARC/CCL17, MDC/CCL22, and Mig/CXCL9) are elevated in sera from patients with atopic dermatitis. *Journal of Dermatological Science* **34**, 201–208, <https://doi.org/10.1016/j.jdermsci.2004.01.001> (2004).
38. Claudio, E. *et al.* Cutting Edge: IL-25 Targets Dendritic Cells To Attract IL-9-Producing T Cells in Acute Allergic Lung Inflammation. *Journal of Immunology (Baltimore, Md.: 1950)* **195**, 3525–3529, <https://doi.org/10.4049/jimmunol.1500436> (2015).
39. Sudlow, C. *et al.* UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Medicine* **12**, e1001779, <https://doi.org/10.1371/journal.pmed.1001779> (2015).
40. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv*, 166298, <https://doi.org/10.1101/166298> (2017).
41. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2018).
42. Jeffrey T. L. *et al.* sva: Surrogate Variable Analysis. R package Version 3.30.1, <https://bioconductor.org/packages/release/bioc/html/sva.html> (2019).
43. Schaum, N. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372, <https://doi.org/10.1038/s41586-018-0590-4> (2018).
44. Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174**, 1015–1030.e1016, <https://doi.org/10.1016/j.cell.2018.07.028> (2018).
45. Hatch, S. L. *et al.* Identifying socio-demographic and socioeconomic determinants of health inequalities in a diverse London community: the South East London Community Health (SELCoH) study. *BMC Public Health* **11**, 861, <https://doi.org/10.1186/1471-2458-11-861> (2011).
46. Hatch, S. L. *et al.* Discrimination and common mental disorder among migrant and ethnic groups: findings from a South East London Community sample. *Social Psychiatry and Psychiatric Epidemiology* **51**, 689–701, <https://doi.org/10.1007/s00127-016-1191-x> (2016).
47. Pamos, A. B. *et al.* Genetic Risk for Psychiatric Disorders and Telomere Length. *Front Genet* **9**, 468, <https://doi.org/10.3389/fgene.2018.00468> (2018).
48. Pamos, A. B. *et al.* Associations between childhood maltreatment and inflammatory markers. *BJPsych Open* **5**, e3, <https://doi.org/10.1192/bjo.2018.80> (2019).
49. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, <https://doi.org/10.1186/s13742-015-0047-8> (2015).
50. Zhang, J. P. *et al.* Schizophrenia Polygenic Risk Score as a Predictor of Antipsychotic Efficacy in First-Episode Psychosis. *The American Journal of Psychiatry* **176**, 21–28, <https://doi.org/10.1176/appi.ajp.2018.17121363> (2019).
51. Machiela, M. J. & Chanock, S. J. LDassoc: an online tool for interactively exploring genome-wide association study results and prioritizing variants for functional investigation. *Bioinformatics* **34**, 887–889, <https://doi.org/10.1093/bioinformatics/btx561> (2018).

## Acknowledgements

T.R.P. is funded by a Medical Research Council Skills Development Fellowship (MR/N014863/1). R.R.R.D. received an NIHR Maudsley Biomedical Research Centre Career Development Award, and funding from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brazil, grant BEX 1279/13-0). This study represents independent research part funded by the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the UK Department of Health. D.F.N. and M.M.R. were funded in part by a grant from the NCI R01 CA206488. We thank Dr. Alan Greenberg for helpful comments on the manuscript, and Dr. Paul McLaren and colleagues for providing us with the GWAS summary statistics.

## Author contributions

Designed experiments: T.R.P., D.F.N. Performed the wet-lab experiments: T.R.P. Analyzed the data: T.R.P., R.R.R.D. Contributed knowledge, reagents, their time, revised the manuscript: M.H., S.H., M.M.R., G.D.B., C.M.L. Wrote the paper: T.R.P., R.R.R.D., D.F.N.

## Competing interests

G.D.B. declares receiving funding from Eli Lilly. The other authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-59256-0>.

**Correspondence** and requests for materials should be addressed to T.R.P. or D.F.N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020